A constructional approach to Spanish idiolectal variation: the case of epistemic stance markers and discourse markers

Andrea Mojedano Batel, Aston University

Rationale

- Forensic authorship analysis needs to be explicable and informed by an understanding of language variation
- The idea of authorship attribution is based on two assumptions:
    - that every language user has a unique linguistic style, or idiolect → empirical evidence is needed
    - that features characteristic of that style will recur with a relatively stable frequency → methods need to be able to capture stylistic similarities across genres and modes
- Mojedano et al. (in preparation) → epistemic stance markers and discourse markers (DMs) are idiolectal features that tend to survive genre and mode effects in Spanish

Research questions

- Are epistemic stance markers and discourse markers distinctive of a person's idiolectal style?
- Do epistemic stance markers and discourse markers remain stable across time?
- Aim → Explore the potential use of these features in authorship analysis tasks

Theoretical framework: construction grammar

- Relationships between patterns within a paradigm are part of a speaker's knowledge → the use of one particular item or constructional variant is understood as related to the functioning of other items related to it (Enghels, 2018)
- Within networks of related constructional variants, new patterns can arise (by becoming more frequent and invading the domains of another pattern) while older ones can disappear (Enghels, 2018)
- Instances of a particular network can be seen as a "cloud of exemplars" (Bybee, 2010)
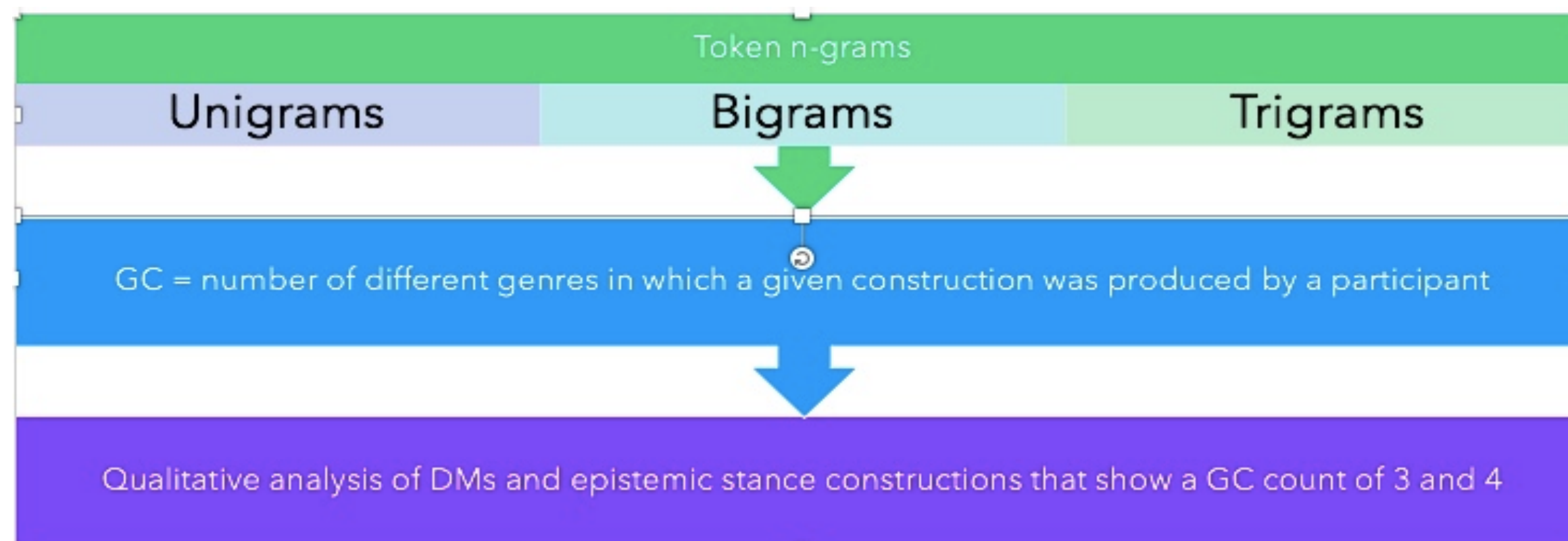
Data

- 9 Spanish-speaking participants from Mexico: 6 women and 3 men
- One community of practice: academic faculty & researchers living in Mexico City, age range: 30-60 y/o
- Multimodal communication: spoken and written registers (formal and informal) → computer and cell-phone keyboard
- Cross-genre perspective: work emails, text messages to family and friends, sociolinguistic interview, work meetings

Genre	Mode of production	Context	Audience
WhatsApp messages (30 per participant, 2020)	Written (phone keyboard)	Informal	Socially close
Emails (30 per participant, 2019-2020)	Written (full-sized keyboard)	Formal	Socially semi-distant
Sociolinguistic interviews (1 per participant, 2020)	Spoken (one-to-one interaction)	Informal	Socially semi-distant
Work meetings (monthly, 2008-2019)	Spoken (group interaction)	Formal	Socially distant

Data: corpus size

	Text messages Token count	Emails Token count	Sociolinguistic interviews Token count	Work meetings Token count	Total
Speaker 1 - female	326	1179	2919	65858	70279
Speaker 2 - female	503	6549	2148	81560	90760
Speaker 3 - male	270	3185	4259	34297	42011
Speaker 4 - female	2114	5306	1562	19980	28962
Speaker 5 - female	856	4916	956	1266	7994
Speaker 6 - male	397	3114	1104	429681	434296
Speaker 7 - male	332	1562	1387	146647	149928
Speaker 8 - male	636	6388	3110	47056	57190
Speaker 9 - female	688	1004	1397	17759	20848
Total	6122	33203	18839	844104	902268

Data analysis

Token n-grams: Unigrams — Bigrams — Trigrams

GC = number of different genres in which a given construction was produced by a participant

Qualitative analysis of DMs and epistemic stance constructions that show a GC count of 3 and 4

Are features that display cross-genre and/or cross-modal stability individuating?

To answer this question, there is a one-vs-all ranking where one speaker's GC is compared with the mean GC calculated for the other eight speakers in the corpus.

N-gram type	N-gram	Discourse types	DTC	Tokens
2-0	{creer,que}	{{chat,2/503},{meeting,176/81560},{email,10/6549},{interview,3/2148}}	4	191

N-gram type	N-gram	DTCs	Mean DTC	Tokens
2-0	{creer,que}	[['idio100_sp_6', '4'], ['idio100_sp_4', '4'], ['idio100_sp_3', '3'], ['idio100_sp_7', '4'], ['idio100_sp_5', '3'], ['idio100_sp_9', '3'], ['idio100_sp_1', '2'], ['idio100_sp_8', '3']]	3.25	1377

The two features below are not individuating due to their high GC mean; the rest of the features are, as they have a lower GC mean

- Doxastic *creer que* 'to believe that' GC mean = 3.25
- Discourse marker *bueno* 'well' GC mean = 2.38

Results

5 speakers (55.5%, 3 female, 2 male) show diachronically consistent use of DMs and/or stance markers across genres and modes
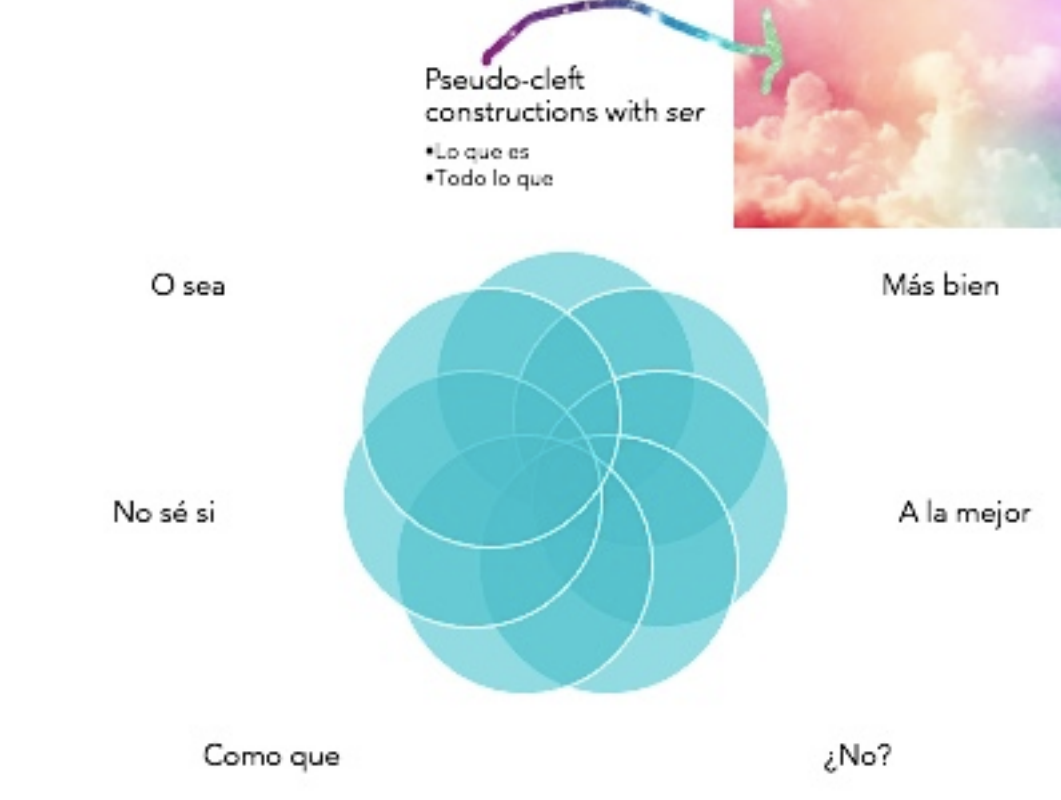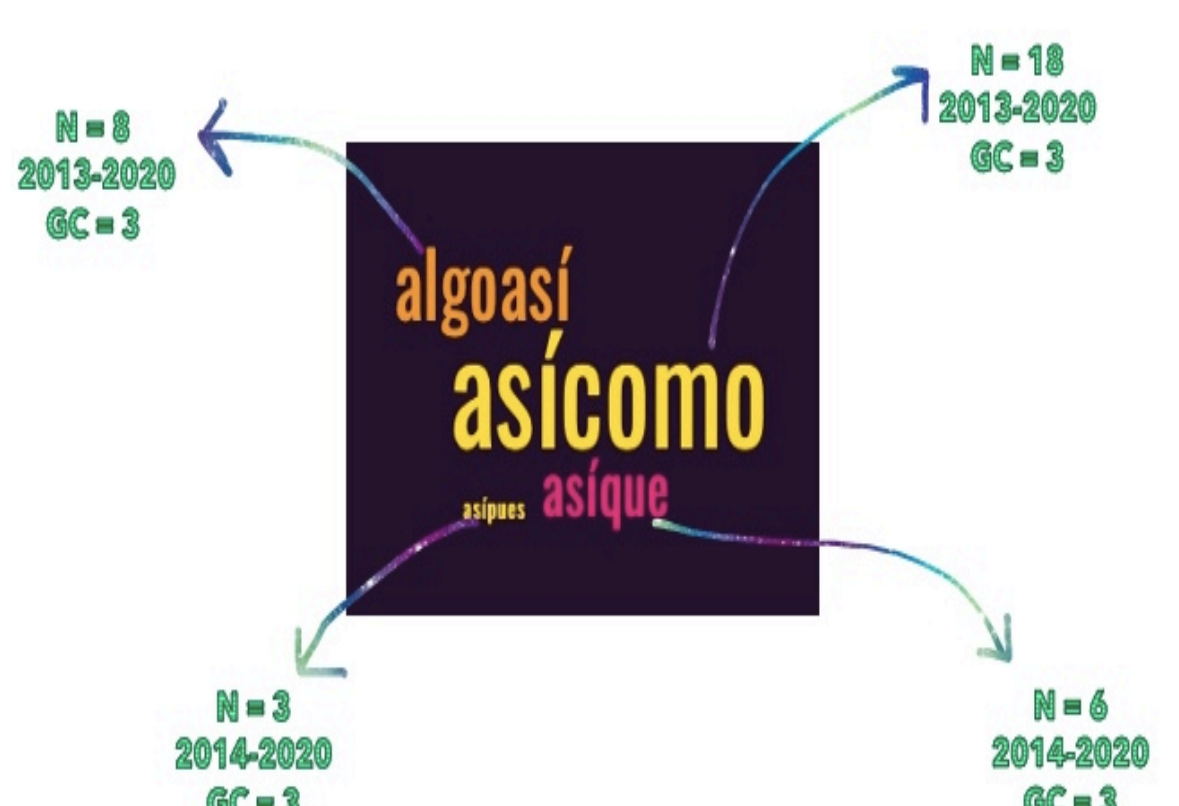
Speaker 2

Speaker 4

Speaker 8

Example: discourse marker *así* ('like this') cloud of exemplars

Results: constructional networks

- A closer look at the *verdad* 'truth' constructional network for Speaker 2 allows us to gauge how this participant uses different schemas in the network for different functions
- *La verdad (es que)* 'the truth (is that)' is an epistemic stance marker (Soler Bonafont, 2016) which displays two formal variants:
    - a parenthetical one, *la verdad*
    - one which is syntactically integrated into the clause it subordinates, *la verdad es que* → reduced version, *la verdad que*
- Speaker 2 produces *la verdad (es que)* in its two variants generally after an adversative (*pero* 'but') or a causal conjunction (*porque* 'because') → attenuating function (González & Maldonado, 2007; Soler, 2015; 2016)
- In turn-initial position, Speaker 2 uses *es verdad que* 'it is true that'

Revisiting the research questions

- Are epistemic stance markers and discourse markers distinctive of a person's idiolectal style?
- → They can be, depending on the frequency of use of these markers in the speech production of other members of the speech community under observation
- Do epistemic stance markers and discourse markers remain stable across time?
- → Preliminary findings point to these features remaining stable for at least 7 years

Preliminary conclusions: language variation in real time

- Results preliminarily point to stance markers and DMs in our data remaining stable for at least 7 years (2013-2020)
- These results help us in partially resolving questions of whether and to what extent speakers change stance markers and discourse markers in adulthood → speakers tend to use the same markers across time, although dispersion has not been examined in this study
- In line with findings that people retain their original patterns throughout their lives, where most aspects of the language of a speech community will not be involved in ongoing change (Sankoff, 2008)

Preliminary conclusions: taking a constructional approach

- The perception of relationships among patterns in a paradigm is part of a speaker's knowledge (Enghels, 2018) → we cannot expect a speaker to necessarily produce only one constructional variant of a schema
- A lexical item such as *así* should not be analyzed as an isolated sequence, but as a part of a wider taxonomy (e.g., *así que*, *así como*, *así pues*)
- Idiolectal study of stance markers and DMs should analyze the frequency and contexts of use of both a constructional schema (e.g., *así* + modifier) and of its constructional variants (e.g., *así que*, *así como*, etc.)
- We should analyze the position of the specific stance expression or DM, which can be related to its pragmatic function on a quantifiable basis (Enghels, 2018) → a form-function correlation could shed light on idiolectal styles, potentially aiding authorship analysis tasks

References

- Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Enghels, R. (2018). Towards a constructional approach to discourse-level phenomena: The case of the Spanish interpersonal epistemic stance construction. *Folia Linguistica, 52*(1), 107-138.
- González, M. & Maldonado, R. (2007). Extensiones pragmáticas de la contraexpectación. Balance, reformulación y réplica. En I. Ibarretxe-Antuñano, C. Inchaurralde & J. Sánchez (Eds.), *Language, Mind and the Lexicon* (pp. 123-142). Frankfurt: Peter Lang.
- Mojedano Batel, A., Alberich Buera, N., & Kredens, K. (in preparation). La variación idiolectal a través de cuatro géneros multimodales: aplicaciones a la atribución de autoría. Aston University.
- Sankoff, G. (2008). Cross-Sectional and Longitudinal Studies. In U. Ammon, N. Dittmar, K. Mattheier & P. Trudgill (Ed.), *Volume 2: An International Handbook of the Science of Language and Society* (pp. 1003-1013). Berlin • New York: De Gruyter Mouton.
- Soler Bonafont, M. A. (2017). La verdad (es que): significado nuclear y atenuante. *Signos. Estudios de Lingüística* 50(95), 430-452.